ON A GENERALIZED METRIC DIMENSION WITH PARTIALLY KNOWN GRAPH TOPOLOGY

Sabina Zejnilovic, Dieter Mitsche, Joao Gomes, Bruno Sinopoli

ON A GENERALIZED METRIC DIMENSION WITH PARTIALLY KNOWN GRAPH TOPOLOGY

SABINA ZEJNILOVIĆ[1,2], DIETER MITSCHE[3], JOÃO GOMES[1], AND BRUNO SINOPOLI[2]

Abstract. The metric dimension of a connected graph G is the minimum number of vertices in a subset S of the vertex set of G such that all other vertices are uniquely determined by their distances to the vertices in S. We introduce and analyze the concept of generalized metric dimension of a disconnected graph, which corresponds to the minimum number of vertices in a subset S such that all other vertices have unique distances to it in all connected graphs that result from completing a given disconnected graph. This generalization allows for incomplete knowledge of the underlying graph in applications such as identifying sources of infection. We quantify the generalized metric dimension exactly when the disconnected components are trees, cycles, grids, complete graphs and give general upper bounds on this number in terms of the boundary of the graph.

1. Introduction

Let G be a finite, simple, connected graph with $|V(G)| = n$ vertices. [1] For a subset $R \subseteq V(G)$ with $|R| = r$, and a vertex $v \in V(G)$, define $\boldsymbol{d}(v, R)$ to be the r-dimensional vector whose i-th coordinate $d(v, R)_i$ is the length of the shortest path between v and the i-th vertex of R. We call a set $R \subseteq V(G)$ a *resolving set* if for any pair of vertices $v, w \in V(G)$, $\boldsymbol{d}(v, R) \neq \boldsymbol{d}(w, R)$. Clearly, the entire vertex set $V(G)$ is always a resolving set, and so is $R = V(G) \setminus \{v\}$ for every vertex v. The *metric dimension* $\beta(G)$ is then the smallest cardinality of a resolving set. We have the trivial inequalities $1 \leq \beta(G) \leq n-1$, with the lower bound attained for a path, and the upper bound for the complete graph. The metric dimension was introduced by Slater [1] in the mid-1970s, and by Harary and Melter [2]. As a start, Slater [1] determined the metric dimension

[1]Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Avenida Rovisco Pais 1, 1049-001 Lisbon, Portugal

[2]Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States

[3]Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice cedex 02

E-mail addresses: sabinaz@cmu.edu, dmitsche@unice.fr, jpg@isr.ist.utl.pt, brunos@ece.cmu.edu.

1991 *Mathematics Subject Classification.* 05C12, 68R10, 94C15.

Key words and phrases. graph theory, metric dimension.

This research was partially supported by Fundação para a Ciência e a Tecnologia (project PEst-OE/EEI/LA0009/2013, a PhD grant from the Carnegie Mellon-Portugal program) and EU FP7 project MORPH (grant agreement no. 288704).

[1]A conference version of this manuscript will appear in the Proceedings of the engineering conference "IEEE Global Conference on Signal and Information Processing GlobalSIP 2014".

of trees. Two decades later, Khuller, Raghavachari and Rosenfeld [3] gave a linear-time algorithm for computing the metric dimension of a tree, and characterized the graphs with metric dimensions 1 and 2. The metric dimension for many graph classes is known, including random graphs [4], and its calculation has also been extensively studied from a computational complexity point of view (see [5, 3, 6]).

In this paper we introduce and analyze the concept of *generalized metric dimension*: we are given a finite, simple, disconnected graph $F = (V, E)$ with $|V| = n$ consisting of k connected components, denoted by C_i, for $i = 1, \dots k$. Denote the class $\mathcal{H}(F)$ to be the class of all possible connected graphs that can be constructed by adding $k-1$ edges. For a graph $H_1 \in \mathcal{H}(F)$, a vertex $u \in V$ and a set $O \subseteq V$, denote by $\boldsymbol{d}_{H_1}(u, O)$ the distance vector of u to the set O in the graph H_1, that is, $(d_{H_1}(u, O))_i$ is the length of the shortest path between u and the i-th vertex of O in the graph H_1. By a generalized resolving set of a disconnected graph F, we denote a set of vertices O such that for any two different vertices u and v, and any two graphs $H_1, H_2 \in \mathcal{H}(F)$, $\boldsymbol{d}_{H_1}(u, O) \neq \boldsymbol{d}_{H_2}(v, O)$. Denote by $\gamma(F)$, the so called *generalized metric dimension*, the cardinality of a smallest generalized resolving set of a graph F. Note that $\max_{H_i \in \mathcal{H}(F)} \beta(H_i) \leq \gamma(F) \leq n - 1$.

Motivation. The introduction of resolving set by Slater [1] was motivated by the application of placement of a minimum number of sonar detectors in a network, while Khuller, Raghavachari and Rosenfeld [3] were interested in finding the minimum number of landmarks needed for robot navigation on a graph. Recently, the problem of finding the minimum number of agents whose infection times need to be observed in order to identify the first infected agent for a simplified diffusion model was cast as finding the metric dimension of the graph [7]. Similarly, to identify a rumor source in a network based on the times when the nodes first heard a rumor, observed nodes should form a resolving set.

However, in many practical applications, the network topology is only partially known. Often, local connections within communities are well known, while the connections between them are not always observed. This may happen when diseases spread from one community to another through random contact, rather than a known friendship connection, or when novel information is spread through weak, rather than strong, social ties. Hence, the problem of finding the minimum number of network devices or agents needs to be considered for scenarios where not all the edges of the graph are known. We model this incomplete knowledge by assuming that the graph of interest is disconnected, with k components and $k-1$ unobserved edges connecting the components, and we consequently introduce the concept of generalized metric dimension. In order to identify the source of infection or a rumor in such a setting, the group of agents that needs to be observed should form a generalized resolving set. Since the resources for observations are often limited, finding the smallest such group of agents, or equivalently, the generalized metric dimension of the graph, becomes a problem of interest.

Notation. For a connected graph G, $i, j \in V(G)$, denote an $i-j$-path to be a sequence of all different vertices $v_0 = i, v_1, \dots, v_\ell = j$, such that for $i = 0, \dots, \ell - 1$, $\{v_i, v_{i+1}\} \in E(G)$. Let $L(C_i)$ denote the set of all leaves of component C_i, and $K(C_i)$

the set of vertices of degree 3 or more that are connected by paths to one or more leaves, when C_i is a tree. For a fixed component C_j of F, denote by S_j a minimum cardinality resolving set of C_j (so that $\beta(C_j) = |S_j|$.) The $M \times N$-grid with $M, N \geq 2$, is the graph whose vertices correspond to the points in the plane with integer coordinates, x-coordinates being in the range $0, \ldots, M-1$, y-coordinates in the range $0, \ldots, N-1$, and two vertices are connected by an edge whenever the corresponding points are at Euclidean distance 1. The four vertices of degree two are called corner vertices.

For a connected graph G, a vertex v is a *boundary vertex* of u if $d_G(w, u) \leq d_G(v, u)$, for all w that are neighbors of v [8]. A vertex v is a boundary vertex of G if it is a boundary vertex of some vertex of G. The set of all boundary vertices of a vertex u is denoted as $\partial(u)$. The *boundary* of graph G, $\partial(G)$, is the set of all boundary vertices of G. It is well known that the boundary is a resolving set, see [9]. For example, the boundary of a tree is the set of its leaves, whereas the boundary of a grid is the set of its 4 corner vertices, and the boundary of a cycle is the whole vertex set [9].

Statements of results. We state the main results of this paper which are then proved in the following sections.

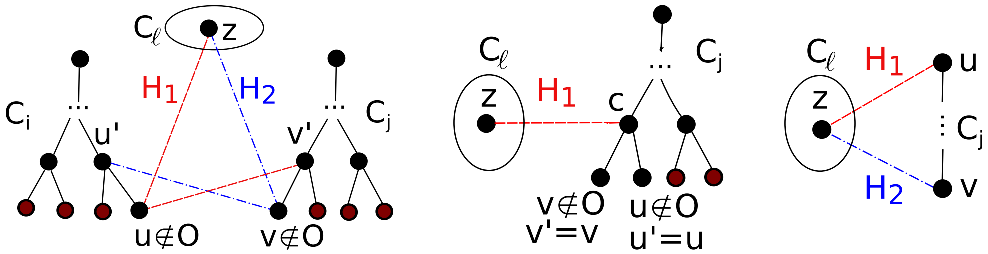
Theorem 1.1. *Let F be a graph of k components, where each component is a tree. Then $\gamma(F) = \min_j \sum_{i=1, i\neq j}^{k} |L(C_i)| + |S_j|$, unless all components are isolated vertices, in which case $\gamma(F) = k-1$. In the first case, we may assume without loss of generality, that the minimum is attained for $j = k$. Then the set consisting of all leaves from components $1, \ldots, k-1$ together with a minimum cardinality resolving set of the k-th component is a minimum cardinality generalized resolving set of the graph F.*

Theorem 1.2. *Let F be a graph of k components, where each component is a complete graph of at least 3 vertices. Then $\gamma(F) = n - k$. A set consisting of all but one vertex of each component is a minimum cardinality generalized resolving set of the graph F.*

Theorem 1.3. *Let F be a graph of k components, where each component is a grid. Then $\gamma(F) = 3k - 1$. Let $O_i = \{r_1^i, r_2^i, r_3^i\}$ denote a set of three corner vertices from component C_i. Then $O = \cup_{i=1}^{k-1} O_i \cup S_k$ is a minimum cardinality generalized resolving set of F.*

Theorem 1.4. *Let F be a graph of k components, where each component is a cycle of size greater than 3. Let k_e denote the number of components with an even number of vertices. Then $\gamma(F) = 2k + k_e - 1$, if $k_e > 0$, and $\gamma(F) = 2k$, otherwise. For a component C_i with an even number of vertices n_i, define $O_i = \{r_1^i, r_2^i, r_3^i\}$, where r_1^i, r_2^i are two neighboring vertices in C_i and r_3^i is a vertex at distance at least $\frac{n_i-2}{2}$ from both of them, also in C_i. For a component C_i with an odd number of vertices n_i, define $O_i = \{r_1^i, r_2^i\}$, where r_1^i and r_2^i are two vertices of C_i that are at distance $\frac{n_i-1}{2}$ from each other. If $k_e = 0$, $\cup_{i=1}^{k} O_k$ is a minimum cardinality generalized resolving set of F. If $k_e > 0$, assume without loss of generality that C_k is a component with an even number of vertices. Then $O = \cup_{i=1}^{k-1} O_i \cup S_k$ is a minimum cardinality generalized resolving set of F.*

For general graph classes we have the following results, the second one tightening the first one, as the boundary of a graph can be very large.



(A) Both components have a leaf which is not included in O.

(B) All leaves of C_i are in O and C_j is not a path.

(C) All leaves of C_i are in O and C_j is a path.

FIGURE 1. Case I in the Proof of Theorem 1.1: Constructing trees H_1 and H_2 when both components C_i and C_j have at least two nodes.

Theorem 1.5. *For any arbitrary graph F with k connected components, the set $O = \cup_{i=1}^{k-1} \partial(C_i) \cup S_k$ is a generalized resolving set for F.*

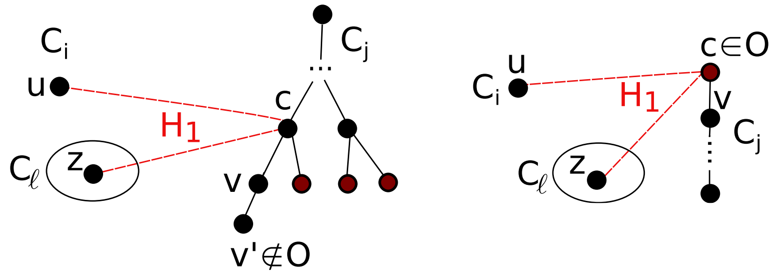
Theorem 1.6. *Let F be an arbitrary graph with k connected components, let $\partial(S_i)$ denote the boundary of the resolving set S_i, and let $O_i = S_i \cup \partial(S_i)$. Then $O = \cup_{i=1}^{k-1} O_i \cup S_k$ is a generalized resolving set for F.*

2. PROOFS OF MAIN RESULTS FOR SPECIAL GRAPH CLASSES

Proof of Theorem 1.1. We first prove the claim of sufficiency. If both u and v are any two vertices in the same component, then u and v are distinguishable as the set of all the leaves of a tree is a resolving set. Hence we may assume $u \in V(C_i)$ and $v \in V(C_j)$ for $i \neq j$. Let p be a vertex in C_i and q a vertex in C_j, such that any path from a vertex in C_i to any vertex in C_j in H_2 contains the subpath $p - q$. Note that $d_{H_2}(p, q) \geq 1$. If u is a leaf, as it is contained in $L(C_i)$, it is distinguishable from v, since $0 = d_{H_1}(u, u) < d_{H_2}(u, v)$. If u is not a leaf, and $u = p$, then for any leaf $r \in L(C_i)$, $d_{H_2}(r, v) = d_{H_2}(r, p) + d_{H_2}(p, q) + d_{H_2}(q, v) \geq d_{H_1}(r, p) + d_{H_2}(p, q) > d_{H_1}(r, p)$. Thus, the two distance vectors are not equal either. Otherwise, if u is not a leaf, and $u \neq p$, let r be a leaf in $L(C_i)$ such that u is on the path from r to p (such a leaf clearly exists). Then $d_{H_2}(r, v) = d_{H_2}(r, u) + d_{H_2}(u, p) + d_{H_2}(p, q) + d_{H_2}(q, v) > d_{H_1}(r, u) + d_{H_1}(u, p) > d_{H_1}(r, u)$. Thus, the two distance vectors also in this case are not equal, which completes the proof of sufficiency.

Now, we prove the claim of necessity. Let O be an arbitrary generalized resolving set. We will show that O has to be at least of the size given by the sufficient condition.

Case I: Let C_i and C_j be two components with at least 2 vertices, such that both have a leaf which is not included in O. Let u be such a leaf in component C_i with neighbor u' and v be a leaf in C_j with neighbor v', such that $u, v \notin O$. We claim that u and v are indistinguishable, as illustrated in Figure 1a. We can construct H_1 by connecting u with v', and u with some vertex z of any other component C_ℓ (if there are more than 2 components). H_2 is then constructed by connecting v with u' and v with the same vertex z as in H_1; the other newly added edges are the same in H_1 and H_2. Now, we have $\boldsymbol{d}_{H_1}(u, O) = \boldsymbol{d}_{H_2}(v, O)$, as follows. For any vertex $r \in C_i \setminus \{u\}$, we have $d_{H_1}(u, r) =$



(A) C_j is not a path. (B) C_j is a path.

FIGURE 2. Case II in the Proof of Theorem 1.1: Constructing trees H_1 and H_2 when component C_i has only one node.

$1 + d_{H_1}(u', r)$, and $d_{H_2}(v, r) = d_{H_2}(u', r) + 1 = d_{H_1}(u', r) + 1$. For any vertex $r \in C_j$, we have $d_{H_1}(u, r) = d_{H_1}(v', r) + 1$, and $d_{H_2}(v, r) = d_{H_2}(v', r) + 1 = d_{H_1}(v', r) + 1$. Finally, for a vertex $r \in C_\ell$, $\ell \neq i, j$, we have $d_{H_1}(u, r) = 1 + d_{H_1}(z, r) = 1 + d_{H_2}(z, r) = d_{H_2}(v, r)$. Thus the vertices u and v are indistinguishable, and the claim holds. Hence, either all the leaves of component C_i or component C_j have to be included in O. Without loss of generality, let us assume that all the leaves of C_i are included in O. Now we assume that only $|S_j| - 1$ vertices are selected from the component C_j. In the first sub-case, when C_j is not a path, from [10], we have $S_j = |L(C_j)| - |K(C_j)|$. If only $|S_j| - 1$ vertices were taken from C_j, then there exists a vertex c in $K(C_j)$ such that two of its associated leaves u and v are both not in O. But then there exists a vertex u' on the path $c - u$, and a vertex v' on the path $c - v$, such that $d_{C_j}(u', c) = d_{C_j}(v', c)$. Note that u' might coincide with u, and v' might coincide with v. The vertices u' and v' are indistinguishable from each other in C_j. Constructing a tree H_1 by connecting any vertex z from any other component C_ℓ, $\ell \neq j$, with any fixed vertex in $K(C_j)$, we see that u' and v' still are indistinguishable by vertices in O, as shown in Figure 1b. In the second sub-case, when C_j is a path with terminal vertices u and v, S_j comprises only one terminal vertex. If neither of the terminal vertices of C_j are in O, H_1 can be constructed by connecting one of its terminal vertices u with any vertex z of any other component C_ℓ, while H_2 is constructed by connecting z to the other terminal vertex v, and vertices u and v are indistinguishable, as Figure 1c shows. Thus, at least $|S_j|$ vertices have to be taken from C_j.

Case II: C_i consists of only one vertex, u, and C_j has more than 2 vertices. With the same arguments as in Case I, it can be seen that at least $|S_j|$ vertices from component C_j have to be included in O. We will show now that u has to be included in O as well. In the first sub-case, when C_j is not a path, then H_1 is constructed by connecting u with a vertex c in $K(C_j)$, and then connecting c to any other component C_ℓ, $\ell \neq i, j$. Let v' be the leaf associated with c, but not in O and let v be a neighbor of c in C_j which lies on the path $c - v'$. Then u is indistinguishable within H_1 from v, as shown in Figure 2a. As for the second sub-case, when C_j is a path, H_1 can be constructed by connecting u with the terminal vertex c of C_j where $c \in O$, and then connecting c to a vertex z of any other component C_ℓ. Let v be a vertex in C_j which is a neighbor of c.

If u is not chosen, u is indistinguishable within H_1 from v, as can be seen in Figure 2b. Hence, u must also be included in O.

Case III: Both C_i and C_j contain only one vertex. Call these u and v, respectively. At least one of them has to be included in O: otherwise, we can construct H_1 by connecting both u and v to some vertex z from any other component C_ℓ, $\ell \neq i, j$, and then u and v are indistinguishable within H_1.

Therefore, for any pair of components C_i and C_j, a generalized resolving set O has to include all leaves from one component and a resolving set from the other, unless both have size 1, in which case only 1 vertex is enough. Hence, if there exists at least one component which has 2 or more vertices, from all but one component all the leaves have to be taken, and from the remaining component, at least a resolving set. If all k components have only one vertex, the set O has to contain $k-1$ vertices. □

Proof of Theorem 1.2. First, we prove the claim of sufficiency. Let us denote the set of all but one vertex on component C_i by O_i. If u and v are in the same component, they are distinguishable, since each O_i is a resolving set of component C_i [3]. Hence, let us assume that vertex $u \in V(C_i)$ is not included in O_i, and that vertex $v \in V(C_j)$ is not included in O_j. Let $p \in V(C_i)$ and $q \in V(C_j)$, such that $p - q$ is the path connecting components C_i and C_j in H_2, so that $d_{H_2}(p, q) \geq 1$. We prove the claim by contradiction and assume that the following relations hold:

$$d_{H_1}(u, r) = 1 = d_{H_2}(v, r) = d_{H_2}(v, q) + d_{H_2}(q, p) + d_{H_2}(p, r),$$

for every $r \in O_i$. Then $d_{H_2}(p, r) = 0$ would have to hold for all $r \in O_i$, which is not possible, and proves the claim.

To prove the claim of necessity, we assume that in one component C_i there are two vertices, u and v, that are not included in O_i. We construct H_1 by adding the edges between a fixed vertex $z \in V(C_i)\backslash\{u, v\}$ and some fixed vertex in each other component. Then we have $d_{H_1}(u, r) = d_{H_1}(v, r)$ for all $r \in O_l$, $l = 1, \ldots, k$, and this completes the proof. The theorem for trees discusses the case when all the components have 1 or 2 vertices. □

Proof of Theorem 1.3. Let us denote the size of the grid C_i as $x_i \times y_i$. We assume that each vertex $l \in V(C_i)$ has assigned to it a position vector (x_l, y_l) which represents its location in the integer lattice C_i, with the first selected corner vertex r_1^i at position $(0, 0)$, r_2^i at $(x_i, 0)$ and r_3^i at $(0, y_i)$. First, let us prove the claim of sufficiency. If u and v are in the same component, they are distinguishable, since any two corner vertices having the same value in one coordinate form a resolving set of a grid [3]. Hence, let us assume that $u \in V(C_i)$ and $v \in V(C_j)$, for $i \neq j$ and $i < k$. Let p be a vertex in C_i and q a vertex in C_j, such that pq is the edge that connects components C_i and C_j in H_2, with $d_{H_2}(p, q) \geq 1$. If $u = p$, then for all $r \in O_i$ we have $d_{H_2}(v, r) = d_{H_2}(r, p) + d_{H_2}(p, q) + d_{H_2}(q, v) > d_{H_2}(r, p) = d_{H_1}(r, u)$. Therefore u and v are distinguishable. For $u \neq p$, let us prove the claim by contradiction. Assuming

$\boldsymbol{d}_{H_1}(u, O_i) = \boldsymbol{d}_{H_2}(v, O_i)$, we obtain the following equations:

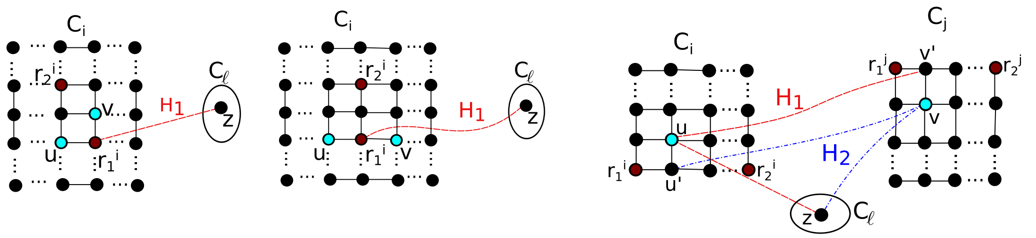
$$\begin{aligned} &d_{H_1}(u, r_1^i) = x_u + y_u \\ =\ &d_{H_2}(v, r_1^i) = x_p + y_p + d_{H_2}(p, q) + d_{H_2}(q, v) \\ &d_{H_1}(u, r_2^i) = x_i - x_u + y_u \\ =\ &d_{H_2}(v, r_2^i) = x_i - x_p + y_p + d_{H_2}(p, q) + d_{H_2}(q, v) \\ &d_{H_1}(u, r_3^i) = x_u + y_i - y_u \\ =\ &d_{H_2}(v, r_3^i) = x_p + y_i - y_p + d_{H_2}(p, q) + d_{H_2}(q, v). \end{aligned} \tag{1}$$

The system of equations (1) has a single solution $x_u = x_p$ and $y_u = y_p$, and $d_{H_2}(p, q) + d_{H_2}(q, v) = 0$, contradicting $d_{H_2}(p, q) \geq 1$. The set $\cup_{i=1}^{k-1} O_i \cup S_k$ is a set of cardinality $3k - 1$, and this completes the sufficiency claim.

For the claim of necessity, let us assume that there exist two components C_i and C_j, such that from each of them, only two vertices are chosen. Let $\{r_1^i, r_2^i\}$ be the set of two vertices from C_i and let $\{r_1^j, r_2^j\}$ be the set of two vertices from C_j that are included in O.

Case I: In at least one component, the vertices included in O are not two corner vertices with one identical coordinate. Let us assume that this is the case with C_i. We claim that there exist two vertices u and v in C_i which are indistinguishable by r_1^i and r_2^i. Denote by $(x_{r_1^i}, y_{r_1^i})$ and by $(x_{r_2^i}, y_{r_2^i})$ the positions at which r_1^i and r_2^i are located in the grid. First, let us consider the sub-case when r_1^i and r_2^i differ in both coordinates, as shown in Figure 3a. Without loss of generality, let us assume that $y_{r_1^i} < y_{r_2^i}$. Then let u be a vertex at $(x_{r_2^i}, y_{r_1^i})$ and v be a vertex at position $(x_{r_1^i}, y_{r_1^i} + |x_{r_2^i} - x_{r_1^i}|)$. Now we have $d_{C_i}(u, r_1^i) = |x_{r_2^i} - x_{r_1^i}| = d_{C_i}(v, r_1^i)$ and $d_{C_i}(u, r_2^i) = y_{r_2^i} - y_{r_1^i} = d_{C_i}(v, r_2^i)$, and hence the vertices u and v are indistinguishable. In the second sub-case, r_1^i and r_2^i differ in only one coordinate, as Figure 3b illustrates. Then, let u and v be two neighbors of r_1^i, which are not on the shortest path $r_1^i - r_2^i$. These two vertices exist, as all vertices on the grid, except the corner vertices, have at least 3 neighbors. Now, we have $d_{C_i}(u, r_1^i) = 1 = d_{C_i}(v, r_1^i)$ and $d_{C_i}(u, r_2^i) = 1 + d_{C_i}(r_1^i, r_2^i) = d_{C_i}(v, r_2^i)$. Therefore, there always exist two vertices u and v, such that they are not distinguishable by any two vertices of C_i which are not two corner vertices with one identical coordinate. Constructing a tree H_1 by connecting any vertex z from any other component C_ℓ, $\ell \neq i$, with either r_1^i or r_2^i, we see that u and v still are indistinguishable by any vertex in O.

Case II: From both components C_i and C_j, two corner vertices with one identical coordinate are included in O. Let u' be a vertex on C_i that is a neighbor of r_1^i such that it shares one coordinate with both r_1^i and r_2^i. Then let u be a neighbor of u' such that it does not share any coordinates with u'. Similarly, let v' be a vertex in C_j that is a neighbor of r_1^j such that is shares one coordinate with both r_1^j and r_2^j. Then let v be a neighbor of v' such that it does not share any coordinates with v'. We can construct H_1 by connecting u with v' and u with any vertex z of any other component (if there are more than 2 components). Then H_2 is constructed by connecting v with u' and v with the same vertex z as in H_1, as shown in Figure 3c. The distances of u and v from



(A) The vertices r_1^i and r_2^i are not corner vertices and they differ in both coordinates.

(B) The vertices r_1^i and r_2^i are not corner vertices and they differ in one coordinate.

(C) The vertices r_1^i and r_2^i are two corner vertices with one identical coordinate.

FIGURE 3. Proof of Theorem 1.3: Constructing H_1 and H_2 when the components are grids.

the vertices in O are

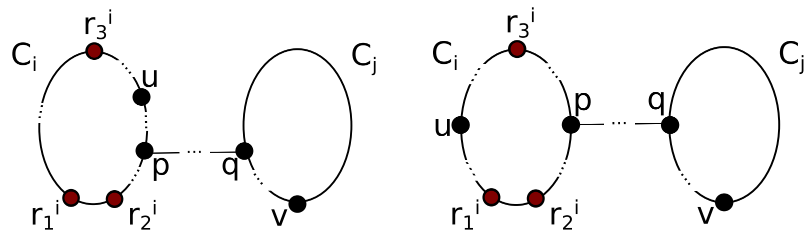
$$\begin{aligned}
d_{H_1}(u, r_1^i) &= d_{H_2}(v, r_1^i) = 2 \\
d_{H_1}(u, r_2^i) &= d_{H_2}(v, r_2^i) = 1 + d_{H_1}(u', r_2^i) \\
d_{H_1}(u, r_1^j) &= d_{H_2}(v, r_1^j) = 2 \\
d_{H_1}(u, r_2^j) &= d_{H_2}(v, r_2^j) = 1 + d_{H_2}(v', r_2^j) \\
d_{H_1}(u, r) \ &= d_{H_2}(v, r) \ = 1 + d_{H_1}(z, r),
\end{aligned}$$

for $r \in C_\ell$, $\ell \neq i, \ell \neq j$. Hence the vertices u and v are indistinguishable.

Therefore, at least 3 vertices of component C_i or component C_j have to be included in O. Without loss of generality, let us assume that 3 vertices in C_i are included in O. Now we assume that only $|S_j| - 1 = 1$ vertices are selected from C_j. Then there exist two vertices u and v in component C_j, which are at the same distance from the only vertex r included from S_j. We construct H_1 by connecting any vertex z from any other component to vertex r in component C_j. Observe that the vertices u and v are still not distinguishable within H_1, and hence at least $|S_j| = 2$ vertices have to be included from component C_j. In conclusion, for any two components, at least 3 vertices from one and 2 vertices from the other one have to be included in O, and thus $|O| \geq 3(k-1) + 2 = 3k - 1$. □

Proof of Theorem 1.4. First, let us prove the claim of sufficiency. As in Theorem 1.3, let us assume that vertex u is located in component C_i and vertex v is in component C_j (when u and v belong to the same component, they are clearly distinguishable, as any two neighboring vertices of an even cycle and any two vertices at distance $(n_i - 1)/2$ in the case of an odd cycle form a resolving set of a cycle). Let components C_i and C_j be connected through the path $p - q$, with $p \in V(C_i)$, and $q \in V(C_j)$. If the vertices u and v are not distinguishable by O_i, then $d_{H_1}(u, r) = d_{H_2}(v, r) = d_{H_1}(p, r) + d_{H_2}(p, q) + d_{H_2}(q, v)$ holds for some H_1 and H_2 and all $r \in O_i$. Therefore,

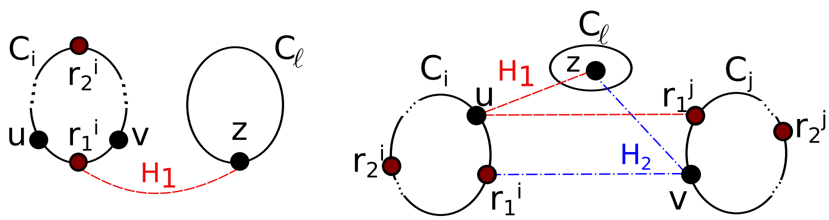
$$d_{H_1}(u, r) > d_{H_1}(p, r) \tag{2}$$



(A) The vertices p and u lie on the same semi-cycle.

(B) The vertices p and u lie on different semi-cycles.

FIGURE 4. Case I in the Proof of Theorem 1.4: Both cycle components have an even number of vertices.



(A) The vertices r_1^i and r_2^i are exactly at distance $\frac{n_i}{2}$.

(B) The vertices r_1^i and r_2^i are not at distance $\frac{n_i}{2}$.

FIGURE 5. Proof of Theorem 1.4: Constructing H_1 and H_2 when the components are cycles.

must hold.

Case I: Both components C_i and C_j have an even number of vertices. Let us first consider the sub-case where both p and u lie in the same half of the cycle, i.e., both lie either on the shorter path $r_2^i - r_3^i$ or on the shorter path $r_1^i - r_3^i$, as shown in Figure 4a. Suppose without loss of generality that they both lie on the shorter path $r_2^i - r_3^i$. As one of the vertices out of $\{u, p\}$ is closer to r_3^i and the other one is closer to r_2^i, (2) cannot hold simultaneously for both r_2^i and r_3^i. The other sub-case that needs to be considered is when u and p lie in different semi-cycles, one on the shorter path $r_2^i - r_3^i$, and the other on the shorter path $r_1^i - r_3^i$, as illustrated in Figure 4b. Then either we have $d_{H_1}(u, r_1^i) = d_{H_1}(u, r_2^i) + 1$ and $d_{H_1}(p, r_1^i) = d_{H_1}(p, r_2^i) - 1$, or $d_{H_1}(u, r_1^i) = d_{H_1}(u, r_2^i) - 1$ and $d_{H_1}(p, r_1^i) = d_{H_1}(p, r_2^i) + 1$. In either case, $d_{H_1}(u, r) = d_{H_2}(v, r)$ cannot hold for both $r = r_1^i$ and $r = r_2^i$.

Case II: At least one of the components C_i or C_j has an odd number of vertices. Let us assume that this is the case with C_i. Similarly, as in Case I, let us first consider the sub-case where both p and u lie in the same half of the cycle, i.e. both on the shorter path $r_1^i - r_2^i$ or both on the longer path $r_1^i - r_2^i$. As before, one of the vertices out of $\{u, p\}$ is closer to r_1^i, and the other is closer to r_2^i, and thus (2) cannot hold simultaneously for both r_1^i and r_2^i. The other sub-case that needs to be considered is when u and p lie in different semi-cycles, one on the shorter path $r_1^i - r_2^i$, of length $\frac{n_i - 1}{2}$, and the other on the longer path $r_1^i - r_2^i$, of length $\frac{n_i + 1}{2}$. Then either we have $d_{H_1}(u, r_2^i) =$

$\frac{n_i-1}{2} - d_{H_1}(u, r_1^i)$ and $d_{H_1}(p, r_2^i) = \frac{n_i+1}{2} - d_{H_1}(p, r_1^i)$, or $d_{H_1}(u, r_2^i) = \frac{n_i+1}{2} - d_{H_1}(u, r_1^i)$ and $d_{H_1}(p, r_2^i) = \frac{n_i-1}{2} - d_{H_1}(p, r_1^i)$. From $d_{H_1}(u, r_2^i) > d_{H_1}(p, r_2^i)$ as given by Condition (2), we obtain $d_{H_1}(p, r_1^i) > d_{H_1}(u, r_1^i) + 1$ or $d_{H_1}(p, r_1^i) > d_{H_1}(u, r_1^i) - 1$. In either case, we get that (2) cannot hold for both $r = r_1^i$ and $r = r_2^i$.

Note that when comparing components C_i and C_j with $i \neq j$, only vertices of the generalized resolving set coming from component C_i were used to distinguish between any two vertices from components C_i and C_j. Hence, for one component, say, C_k, it is enough to choose a resolving set, that is, a set that distinguishes all vertices within C_k (a minimum cardinality resolving set is always of size 2). Hence, if $k_e > 0$, we may assume that C_k is an even cycle. Thus only 2 vertices are chosen from C_k, and from all other even cycles 3 vertices are chosen. Thus, in this case $2k + k_e - 1$ vertices are enough. If $k_e = 0$, then 2 vertices are chosen from each component, giving the bound $2k$ in this case.

Now, we prove the claim of necessity. Observe first that clearly at least 2 vertices of each cycle have to be chosen, as otherwise the two neighbors of the chosen vertex r cannot be separated; one can construct a graph H_1 by connecting r with one fixed vertex of each other component, and the two neighbors of r are indistinguishable.

Let us first assume that there exist two components C_i and C_j both containing an even number of vertices, and from each component, only two vertices are included in O. Denote by r_1^i, r_2^i the vertices chosen from C_i and by r_1^j, r_2^j the vertices chosen from C_j. If in at least one component, say C_i, the two selected vertices r_1^i and r_2^i are at distance exactly $\frac{n_i}{2}$ from each other, let u and v be two neighbors of r_1^i. Note that u and v are equidistant from both r_1^i and r_2^i. Constructing H_1 by connecting any vertex z from any other component C_ℓ to r_1^i, the vertices u and v are still not distinguishable within H_1, as shown in Figure 5a. Otherwise, let us assume that in both components C_i and C_j the vertices selected in O are not at distance exactly $\frac{n_i}{2}$ ($\frac{n_j}{2}$, respectively) from each other. Let u then be a neighbor of r_1^i in C_i that is on the longer path $r_1^i - r_2^i$, and let v be a neighbor of r_1^j in C_j that is on the longer path $r_1^j - r_2^j$. We can construct H_1 by connecting u with r_1^j and u with some vertex z of any other component (if there are more than 2 components). H_2 is constructed by connecting v with r_1^i and v with the same vertex z as in H_1, as shown in Figure 5b. The distances of the vertices u, v from the vertices in O are

$$\begin{aligned}
d_{H_1}(u, r_1^i) &= d_{H_2}(v, r_1^i) = 1 \\
d_{H_1}(u, r_2^i) &= d_{H_2}(v, r_2^i) = 1 + d_{H_1}(r_1^i, r_2^i) \\
d_{H_1}(u, r_1^j) &= d_{H_2}(v, r_1^j) = 1 \\
d_{H_1}(u, r_2^j) &= d_{H_2}(v, r_2^j) = 1 + d_{H_2}(r_1^j, r_2^j) \\
d_{H_1}(u, r) \ &= d_{H_2}(v, r) \ \ = 1 + d_{H_1}(z, r),
\end{aligned}$$

for $r \in O_l$, $l \neq i, j$. Hence the vertices u and v are indistinguishable.

Therefore, if both C_i and C_j have an even number of vertices, at least 3 vertices of component C_i or 3 vertices of component C_j have to be included in O. Hence, from all but one component with an even number of vertices, 3 vertices have to be chosen, and from the remaining ones, at least 2. This completes the proof. □

3. Proofs of results for general graph classes

We start with the following easy observation.

Observation 3.1. *Let G be a connected graph. Consider any two vertices r and u of G, and consider a shortest path $r-u$. Then either u is a boundary vertex for r, or there exists some vertex u' such that the shortest path $r-u$ can be extended to a shortest path $r-u'$, with u' being a boundary vertex for r.*

Proof. If u is not a boundary vertex for r, then by definition there exists a neighbor w of u such that $d_G(w,r) > d_G(u,r)$. Thus, $d_G(w,r) \geq d_G(u,r)+1$, and in particular, a shortest path $r-u$ can be extended to w such that along this extended path, the lower bound can be attained, and thus $d_G(w,r) = d_G(u,r)+1$. Hence, the path $r-w$ going through u is also a shortest path $r-w$. If w is then a boundary vertex for r, we are done, and otherwise we iteratively apply the same argument with w playing the role of u. The claim follows. □

We are now ready to show our results in terms of boundary vertices.

Proof of Theorem 1.5. Since the boundary is a resolving set, any two vertices belonging to the same component are distinguishable by a set that contains the boundaries of $k-1$ component and a resolving set of the k-th component. As before, let $u \in V(C_i)$, $v \in V(C_j)$, let $p \in V(C_i)$ and $q \in V(C_j)$ such that any path from a vertex in C_i to any vertex in C_j in H_2 contains the subpath $p-q$, and let $i<k$. As in the previous theorems, we need to show only the case $u \neq p$. If u is a boundary vertex for p, let $u'=u$. Otherwise, the shortest path between p and u in component C_i can be extended to a shortest path $p-u'$ by Observation 3.1, such that u' is a boundary vertex of p. For a fixed shortest path $p-u'$ we have $d_{H_2}(u',v) = d_{H_2}(u',p)+d_{H_2}(p,q)+d_{H_2}(q,v) = d_{H_1}(u',u)+d_{H_1}(u,p)+d_{H_2}(p,q)+d_{H_2}(q,v) > d_{H_1}(u',u)$, which completes the proof. □

Proof of Theorem 1.6. Let $r \in S_i$ be a vertex from a resolving set of a component C_i. Once more, let $u \in V(C_i)$, $v \in V(C_j)$, $p \in V(C_i)$, $q \in V(C_j)$ such that any path from a vertex in C_i to any vertex in C_j in H_2 contains the subpath $p-q$, and let $i<k$. As in the proof of Theorem 1.5, if u is a boundary vertex for r, let $u=u'$. Otherwise, by Observation 3.1, the shortest path between r and u in component C_i can be extended to a shortest path $r-u'$, with u' being a boundary vertex for r. We need to show that $d_{H_1}(u,u') \neq d_{H_2}(v,u')$, for any vertex v belonging to some other component C_j (as in the previous theorems, if u and v are in the same component, they are distinguishable by the resolving set of that component). If u is a boundary vertex itself, then we clearly have $d_{H_1}(u,u') = 0 \neq d_{H_2}(v,u')$, so we may assume $u \neq u'$. If r does not distinguish u and v, then $d_{H_1}(u,r) = d_{H_2}(v,r) = d_{H_1}(p,r)+d_{H_2}(p,q)+d_{H_2}(q,v)$ and

$$d_{H_1}(u,r) > d_{H_1}(p,r), \tag{3}$$

holds, since $d_{H_2}(p,q)=1$.

Case I: There exists a shortest path from u' to p, and consequently to v, that passes through u. Then we have $d_{H_2}(u',v) = d_{H_1}(u',u)+d_{H_1}(u,p)+d_{H_2}(p,v) > d_{H_1}(u',u)$. Thus u and v have different distances to u', and they are distinguishable.

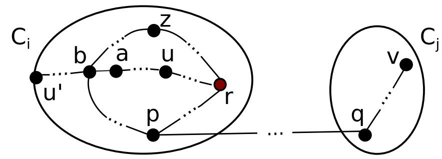


FIGURE 6. Proof of Theorem 1.6: Extending the shortest path $r-u$ to a shortest path $r-u'$.

Case II: All shortest paths from u' to p, and consequently to v, do not pass through u. Let b be the vertex closest to u on this path, such that the path $b-u'$ is common to both paths $p-u'$ and $r-u'$. Note that b might coincide with u', but not with u, as this is already handled by Case I. Next, we claim that at least one shortest path $r-b$ passes through u.

Let us assume the opposite, i.e. there exists a vertex z such that

$$d_{H_1}(r,u)+d_{H_1}(u,b) > d_{H_1}(r,z)+d_{H_1}(z,b). \tag{4}$$

If there are several such z, we pick any vertex z minimizing the right hand side. This vertex z can also be vertex p itself. Now let a be a vertex that immediately precedes b on the (directed) path $u-u'$, as illustrated in Figure 6. Such a vertex exists, as b cannot be vertex u under the assumptions of Case II. Then $d_{H_1}(r,a) = \min\{d_{H_1}(r,z)+d_{H_1}(z,b)+1, d_{H_1}(r,u)+d_{H_1}(u,b)-1\}$. Indeed, $d_{H_1}(r,a)$ cannot be smaller than $d_{H_1}(r,u)+d_{H_1}(u,b)-1$, as otherwise the shortest path $r-u$ could not have been extended to a shortest path $r-b$. Now, if the first value is smaller, we have $d_{H_1}(r,a) > d_{H_1}(r,b)$, which is not possible as the distance from vertex r does not decrease along the extended path $r-u'$. If the second value is smaller or both values are equal, and yet we have that (4) holds, then we have $d_{H_1}(r,u)+d_{H_1}(u,b) = d_{H_1}(r,z)+d_{H_1}(z,b)+1$, and thus $d_{H_1}(r,a) = d_{H_1}(r,b)$. This implies that again the shortest path $r-u$ could be extended only to $r-a$, and not to $r-b$, which contradicts our assumptions and proves the claim.

Since at least one shortest path $r-b$ passes through u, we have

$$d_{H_1}(r,u)+d_{H_1}(u,b) \le d_{H_1}(r,p)+d_{H_1}(p,b). \tag{5}$$

Since (3) holds, from (5) it follows that

$$d_{H_1}(u,b) < d_{H_1}(p,b). \tag{6}$$

Now, $d_{H_2}(v,u') = d_{H_2}(v,p)+d_{H_2}(p,b)+d_{H_2}(b,u') > d_{H_2}(v,p)+d_{H_2}(u,b)+d_{H_2}(b,u') > d_{H_1}(u,u')$. The first inequality follows from (6), and the second inequality uses the fact that $d_{H_2}(v,p) = d_{H_2}(v,q)+d_{H_2}(q,p) \ge 1$. Therefore, u and v have different distances to the boundary vertex u', and they are thus distinguishable by a boundary vertex of a vertex belonging to the resolving set. The theorem follows. □

Remark 3.2. *Inspecting the proofs of Theorems 1.1, 1.3, 1.4, we see that when comparing two vertices from C_i and C_j, in fact only the structure of C_i and of its resolving set matters. Therefore, whenever one of the components of the observed disconnected graph F is a tree, (cycle, or grid, respectively), then instead of including a resolving set and its boundary vertices, it is sufficient to choose all leaves in the case the component* *is a tree (two neighboring vertices together with vertex at distance at least* $\frac{n-2}{2}$ *from both of them in the case of the even cycle on* n *vertices, two vertices at distance* $\frac{n-1}{2}$ *from each other in the case of an odd cycle on* n *vertices, and three corner vertices in the case of the grid, respectively). Note that this might be better than the bound claimed by Theorem 1.6, which for example in the case of the grid requires all four corner points to be chosen.*

4. Concluding remarks

We have introduced and analyzed the concept of a generalized metric dimension for different graph classes. The proposed metric enables the introduction of uncertainty in graph topology in problems modeled with metric dimension. One such problem is to find the minimum number of observed nodes needed for identification of the source node of network diffusion, in the settings where knowing the full network topology is not feasible.

We have given exact answers on this generalized metric dimension for trees, cycles, grids, and complete graphs, and have given general upper bounds for arbitrary graphs in terms of their boundary. Needless to say, it would be interesting to determine this number exactly for other graph classes, such as bipartite graphs, or to find tighter bounds. Additionally, in practical scenarios involving network diffusion, links connecting the vertices of the network represent stochastic propagation times of some rumor or a virus. Hence, it would be of practical interest to analyze a suitably defined stochastic version of both the standard and generalized metric dimension problems.

References

[1] P. Slater, "Leaves of trees," *Congressus Numerantium*, vol. 14, pp. 549–559, 1975.

[2] F. Harary and R. A. Melter, "The metric dimension of a graph," *Ars Combinatoria*, vol. 2, pp. 191–195, 1976.

[3] S. Khuller, B. Raghavachari, and A. Rosenfeld, "Landmarks in graphs," *Discrete Applied Mathematics*, vol. 70, no. 3, pp. 217–229, 1996.

[4] B. Bollobás, D. Mitsche, and P. Prałat, "Metric dimension for random graphs," *The Electronic Journal of Combinatorics*, vol. 20, no. 4, p. P1, 2013.

[5] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, 1979.

[6] J. Díaz, O. Pottonen, M. J. Serna, and E. J. van Leeuwen, "On the complexity of metric dimension," in *Algorithms – ESA 2012*, ser. Lecture Notes in Computer Science, vol. 7501. Springer Berlin Heidelberg, 2012, pp. 419–430.

[7] S. Zejnilovic, J. Gomes, and B. Sinopoli, "Network observability and localization of the source of diffusion based on a subset of nodes," *51st Annual Allerton Conference on Communication, Control, and Computing*, pp. 847– 852, 2013.

[8] G. Chartrand, D. Erwin, G. L. Johns, and P. Zhang, "Boundary vertices in graphs," *Discrete Mathematics*, vol. 263, pp. 25–34, 2003.

[9] C. Hernando, M. Mora, I. M. Pelayo, and C. Seara, "Some structural, metric and convex properties on the boundary of a graph," *Electronic notes in Discrete Mathematics*, vol. 24, pp. 203–209, 2006.

[10] G. Chartrand, L. Eroh, M. A. Johnson, and O. R. Oellermann, "Resolvability in graphs and the metric dimension of a graph," *Discrete Applied Mathematics*, vol. 105, pp. 99–113, 2000.